FRICATIVE PERCEPTION BY COCHLEAR IMPLANT USERS

P.J. Blamey* and V.C. Tartter+

* Human Communication Research Centre
University of Melbourne

+ Department of Psychology
City College, City University of New York

ABSTRACT - Three implant users were tested with 45 syllables consisting of [v,f,ð,θ,z,s, ʒ,ʃ,dʒ,tʃ,h,t,d,n,l] before the vowels [i,a,u] with three wearable speech processors. The WSP3 processor coded first and second formant frequencies and amplitudes. The MSP1 processor used a similar scheme with improved measurement and coding of the formants. The MSP2 processor added amplitude information from three higher frequency bands. Average scores were 42% for WSP3, 54% for MSP1, and 57% for MSP2. Perception of voicing, manner, and place of articulation of the consonants was significantly greater for the MSP processors than the WSP3 processor. Place perception was slightly higher for MSP2 than MSP1. The listeners used three perceptual dimensions which were highly correlated with the frequencies and amplitudes of peaks in the low frequency region of the frication spectrum, amplitudes of high frequency peaks, and duration of the frication noise.

INTRODUCTION

Cochlear implants use electrical stimulation of residual nerve fibres to produce hearing sensations in deaf people. Most implantees now use the 22-electrode implant manufactured by Cochlear Pty Ltd (Clark *et al*, 1984) which electrically stimulates different parts of the cochlea to represent first and second formant frequencies. One might expect fricatives and affricates to be represented poorly by a formant-coding process because they are often characterized by broad spectral distributions including frequencies outside the usual F1 and F2 ranges (Strevens, 1960). Previous studies of the 12 consonants [p,b,m,f,v,s,z,t,d,n,k,g] also showed that [f] and [v] were recognized less often than the other consonants by implant users (Blamey *et al*, 1987b). This experiment tested perception of fricatives and other consonants through three speech processors (WSP3, MSP1, MSP2). The MSP2 processor encoded amplitudes of three high frequency bands as well as the formant information. These extra amplitudes were expected to improve fricative recognition because they provide a representation of the spectrum in the 2-6 kHz region. More general comparisons of the MSP2 and WSP3 processors have been published by Dowell *et al* (1990) and Skinner *et al* (1991).

SPEECH PROCESSING

The WSP3 (Wearable Speech Processor) was based on analog circuitry that measured the fundamental frequency, F0, first and second formant frequencies, F1 and F2, and amplitudes, A1 and A2, and encoded them using a custom-designed digital encoder chip (Seligman, 1987; Blamey *et al*, 1987a). Briefly, F0 was measured by a zero-crossing detector at the output of a full-wave rectifier and 270 Hz low-pass filter to obtain the speech amplitude envelope. F1 was estimated from the zero-crossings of a bandpass filter from 480 Hz (single pole) to 850 Hz (two poles). A1 was estimated as the averaged peak amplitude of the speech signal after low-pass filtering at 850 Hz. F2 and A2 were estimated using zero-crossings and average peak levels from a high-pass filtered waveform, using 4 poles at 850 Hz and 1000 Hz so that F2 predominated over other formants in most vowel spectra for an average of a large number of speakers. For voiceless consonants, the zero-crossing and peak level measuring circuits operated as for voiced phonemes, but obviously the output parameters could not be interpreted as estimates of fundamental frequency and formant frequencies and amplitudes. Nevertheless, these output parameters provided information about voiceless consonants that could be recognised by implant users. F0 controlled the electrical pulse rate. F1 was used to select an electrode pair in the more apical part of the electrode array. F2 was used to select an electrode pair in the more basal part of the array. A1 and A2 were used to control the current levels of the pulses applied to these two electrode pairs.

The MSP1 processor (Miniature Speech Processor) also implemented an F0F1F2 speech coding scheme, but the details were different from the WSP3 as follows: a) Speech parameters were measured and coded within a single digital processing chip designed for this purpose. b) The amplitude measuring algorithms in the MSP1 gave 7 bits of resolution instead of 5 in the WSP3. c) The pulse duration was varied as well as the electrical current to control loudness. This reduced power consumption and increased maximum stimulation rates. d) The loudness of intermediate sounds was increased relative to the loudest and softest sounds. e) The average minimum amplitude in each frequency band was subtracted from the peak amplitude in that band to reduce the effects of continuous background noise. f) Sensitivity was adjustable continuously instead of selection of one of 5 discrete levels. g) F0 was measured using a peak-picking method to give greater reliability and better performance with music and other non-speech sounds. h) The microphone response was extended from 4 to 6 kHz. i) The rate of stimulation during voiceless sounds depended on the timing of peaks in the waveform below 1100 Hz (shown as F1 pulses in Figure 1) with a minimum period of 2 ms. This resulted in a pseudo-random rate with an average of about 250 Hz, compared with a pseudo-random rate of about 130 Hz in the WSP3. j) The bandpass filters used to estimate F1 and F2 had cutoff frequencies as shown in Figure 1.

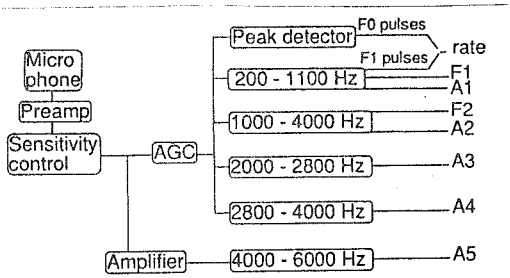

Figure 1. Block diagram of the parameter extraction for the MSP2 speech processor.

The difference between MSP1 and MSP2 was that MSP2 encoded extra speech parameters corresponding to the amplitudes of the speech spectrum in the frequency ranges 2-2.8 kHz, 2.8-4 kHz, and 4-6 kHz. These amplitudes (A3, A4, and A5) were encoded by stimulating three high-pitched electrodes near the basal end of the cochlea. When a sound was voiced, four electrodes were stimulated in each F0 period, corresponding to F1, F2, A3 and A4. For voiceless sounds, no electrode was stimulated for F1 because A1 was normally low, and the A5 electrode was stimulated instead.

SUBJECTS

Three implant users served as subjects. Each listener had been implanted for over 12 months, scored better than average on the post-operative evaluation tests undertaken by all patients implanted in Melbourne, and was available for testing once per week. Psychophysical and speech studies have been reported previously, with the patients identified as 16, 36 and 67 according to their date of operation. Their ages were 63, 47, and 50 years and their lengths of profound deafness prior to implantation were 4, 2, and 2 years respectively.

STIMULI AND PROCEDURE

The listeners were tested with 45 syllables: the fricatives [v,f,ð,θ,z,s,ʒ,ʃ], the affricates [dʒ,tʃ], the aspirate [h], and the additional consonants [t,d,n,l] before the vowels [i,a,u]. The consonants [t,d,n,l] were included to provide information about manner distinctions. Only one vowel context was used within each block of testing, and each syllable occurred three times in random order within the block. The subjects were asked to choose from the closed set of 15 consonants the one that was closest to the one they had just heard. After responding, the subject was informed of the correctness of the response. The patients were tested in weekly sessions and usually, one block of stimuli for

each of the three vowels would be tested within a session. At least six blocks of data for each vowel and each processor were collected for each subject. The stimuli were presented with live voice in a randomized order, spoken by an Australian male who was well known to the three subjects. The live-voice testing was carried out in a sound-treated room using the subjects' own take-home speech processors with input through the directional ear-level microphone as in normal operation. The processor sensitivity was adjusted by the subjects to their preferred value each time. The speaker and subject sat opposite each other at a distance of about 1 m. Lipreading was not used. The order of processor testing was MSP1, MSP2, WSP3 for listener 16; and WSP3, MSP2, MSP1 for listeners 36 and 67. This order of testing was dictated by the usage of the take-home experimental processors used by the subjects since it was necessary to ensure that they had reached a good level of familiarity with each processor before testing was commenced.

RESULTS

Table 1 shows the scores for each subject, each vowel context and each processor. An analysis of variance was carried out in which the results from the separate blocks of stimuli were used as repeated measures, and the independent factors were subject X vowel X processor type. Each main effect was highly significant ($F=153.1$, $df=2,149$, $p<0.001$ for subject; $F=14.5$, $df=2,149$, $p<0.001$ for vowel; $F=52.0$, $df=3,149$, $p<0.001$ for processor type). The only interaction term that was significant was the subject X processor term ($F=12.5$, $df=6,149$, $p<0.001$). The Newman-Keuls post hoc comparison test with $p<0.001$ indicated significant differences between the mean scores for all three subjects (37%, 64%, and 52%). The mean score of 47% for [i] was significantly less than for [u] (52%) and [a] (54%). There were also significant differences between the mean scores for WSP3 (42%) and the two MSP processors (54% and 57%).

Speech processor	Patient 16			Patient 36			Patient 67		
	i	a	u	i	a	u	i	a	u
WSP3	29	37	41	50	50	52	36	47	40
MSP1	34	44	36	60	67	65	54	65	61
MSP2	32	39	43	70	81	76	54	60	56

Table 1. Percentage correct scores.

In order to provide more detailed information about the differences between the processors, 27 confusion matrices were compiled (3 patients X 3 processors X 3 vowel contexts). The percentages of information transmission for the features voicing, place of articulation, and manner of articulation were calculated for each matrix. For each feature, a two-way analysis of variance (patient X processor) was carried out using the three vowel contexts as repeated measures. Each ANOVA showed a highly significant main effect for patient and for processor ($p<.001$). Table 2 shows the average percentage information transmission for each feature and each processor. In the table, the < symbol indicates that the value to the left is significantly less than the value to the right (Newman-Keuls test, $p<0.05$). For every feature, the MSP1 and MSP2 processors produced higher information transmission than the WSP3 processor. MSP2 produced higher information transmission than MSP1 for consonant place, but not for voicing or manner of articulation.

Feature	Processor				
	WSP3		MSP1		MSP2
Voicing	18	<	33		36
Manner	41	<	61		64
Place	34	<	42	<	51
Total	46	<	57	<	62

Table 2. Percentage of information transmitted.

The confusion matrices were also entered into a multidimensional scaling analysis using the individual differences (INDSCAL) model (Schiffman *et al*, 1981). The analysis automatically places the stimuli (consonants) in an N-dimensional "perceptual space" where each dimension may be related to some quality of the stimuli perceived by the subjects. This analysis can be informative if it is possible to identify acoustic parameters that are highly correlated with the dimensions of the perceptual space. The number of dimensions and the coordinates of the stimuli are adjusted to produce an optimum fit to the confusion matrices. Analyses for 2, 3, and 4 dimensions were carried out. The 3 dimensional space was found to fit the data well, with a relatively small improvement when going to 4 dimensions. The three dimensional space was more easily interpreted in terms of the stimulus parameters than either the 2 or 4 dimensional spaces.

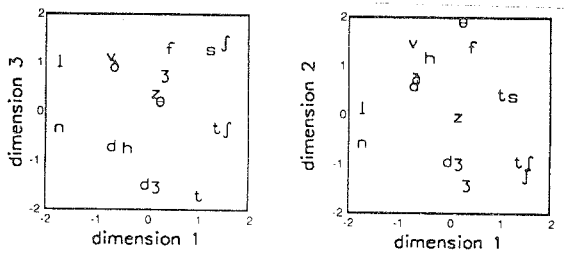

Figure 2. Relative positions of the consonants within the 3-dimensional perceptual space.

Table 3 shows the values of MSP output parameters averaged over the frication portion of the fricatives and affricates, the burst of the two stops, and the initial part of [n] and [l] before the amplitude of the syllable rose to 80% of its maximum value. The A parameters are amplitudes measured on a linear scale. The F1, F2, and rate parameters are in Hz. Rate refers to the electrical pulse rate used, which is equal to F0 for voiced sounds, and a higher pseudo-random rate for unvoiced sounds. The v/uv parameter refers to a voicing decision. The values 0 for unvoiced, 1 for voiced, and 0.5 for partially voiced were assigned for three tokens of each consonant before the vowels [i,a,u] and the values were added. The dur parameter is the duration in ms. Each value in Table 3 (with the exception of v/uv) represents the median value for three tokens of each consonant spoken before the vowels [i,a,u]. The measurements were taken with the standard headset microphone 50 cm from the speaker's mouth and the sensitivity of the processor on 3.

Consonant	rate	v/uv	F1	F2	A1	A2	A3	A4	A5	dur
f	241	0.0	661	2216	5	19	21	30	36	129
s	237	0.0	640	2388	4	13	9	20	148	202
ʃ	257	0.0	864	2525	5	80	101	145	143	167
θ	230	0.0	674	2311	4	12	8	15	27	204
tʃ	245	0.0	787	2320	8	80	120	140	134	107
h	249	0.0	649	1769	17	31	58	27	27	75
v	107	2.5	409	1963	18	29	47	48	46	136
z	185	2.0	435	2164	11	13	10	14	143	137
ʒ	195	1.5	585	2404	12	72	110	139	143	152
ð	256	1.5	446	2034	7	17	26	28	46	75
dʒ	248	0.0	707	2438	9	80	116	136	137	46
t	239	0.0	672	2281	13	38	46	64	80	72
d	192	3.0	486	1718	52	77	149	144	117	19
n	122	3.0	356	1321	29	12	30	11	3	101
l	116	3.0	353	1550	86	46	70	112	80	113

Table 3. Average parameter values for MSP processor.

Table 4 shows correlations between the parameters of Table 3 and the dimensions of the perceptual space. All values greater than 0.50 are significantly different from zero with $p<0.05$ and are shown in bold type. Correlation coefficients with absolute value greater than 0.65 and 0.75 have $p<0.01$ and $p<0.001$ respectively. The parameters fall into three fairly distinct groups, with each group clearly associated with one of the perceptual dimensions. Dimension 1 is associated with voicing and other parameters relating to F1 and F2. Dimension 2 is associated with amplitude in the higher frequency part of the spectrum. Dimension 3 is related only to the duration of the consonant burst or frication.

Parameter	rate	v/uv	F1	F2	A1	A2	A3	A4	A5	dur
rate	**1.00**	**-.85**	**.80**	**.67**	**-.64**	.19	.07	.09	.26	-.02
v/uv	**-.85**	**1.00**	**-.89**	**-.75**	**.71**	-.09	.09	.02	-.16	-.26
F1	**.80**	**-.89**	**1.00**	**.77**	**-.59**	.45	.28	.35	.37	.20
F2	**.67**	**-.75**	**.77**	**1.00**	**-.71**	.29	.08	.25	**.59**	.41
A1	**-.64**	**.71**	**-.59**	**-.71**	**1.00**	.15	.28	.28	-.09	-.35
A2	.19	-.09	.45	.29	.15	**1.00**	**.96**	**.98**	**.61**	-.36
A3	.07	.09	.28	.08	.28	**.96**	**1.00**	**.94**	.50	-.50
A4	.09	.02	.35	.25	.28	**.98**	**.94**	**1.00**	**.62**	-.31
A5	.26	-.16	.37	**.59**	-.09	**.61**	.50	**.62**	**1.00**	.09
dur	-.02	-.26	.20	.41	-.35	-.36	-.50	-.31	.09	**1.00**
dim1	**.73**	**-.80**	**.87**	**.89**	**-.70**	.28	.09	.21	**.57**	.38
dim2	-.05	-.02	-.23	-.22	.00	**-.65**	**-.56**	**-.64**	**-.65**	.09
dim3	-.18	.14	-.14	.11	-.03	-.25	-.32	-.15	.04	**.68**

Table 4. Correlations between acoustic parameters and perceptual dimensions.

DISCUSSION

The largest difference in processors was between WSP3 and MSP1, and was reflected in all three features: voicing, manner, and place of articulation. The improvements in information transmission for voicing and manner were larger than the improvement for place, and probably arose from the differences in the measurement and coding of the A1 and A2 parameters mentioned above in the description of speech processing. It has been noted elsewhere that the details of the amplitude envelope of the speech waveform, and the relative levels of A1 and A2 can provide cues for voicing and manner of articulation (Blamey *et al*, 1985; Van Tasell *et al*, 1987). The only significant difference between MSP1 and MSP2 was for the place feature. Inspection of Tables 3 and 4 shows that the extra parameters A3 and A4 encoded by the MSP2 processor were highly redundant with A2 for this consonant set. A5 was only partly redundant with A2 and could have provided a strong cue to differentiate [s] from [f,ɵ,h] and [z] from [v,ð].

Voicing was the feature with the lowest proportion of information transmitted for all three processors. Inspection of the v/uv column in Table 3 shows that this was possibly caused by errors in the hardware parameter extraction for the voiced fricatives and affricate. If this were so, it would be expected that most of the voicing errors should be in the direction of voiceless responses to voiced stimuli. The numbers of voicing errors for voiced and voiceless stimuli were 403 and 353 for WSP3, 254 and 229 for MSP1, and 305 and 211 for MSP2, consistent with this interpretation.

The multidimensional scaling results suggest that the confusions among these 15 consonants may be explained in terms of three major perceptual dimensions. The first of these was associated with the F1 and F2 frequencies, and with the parameters that reflect voicing: rate, v/uv, and A1. The second dimension was strongly associated with the higher frequency amplitudes A2-A5. The third dimension was associated only with the consonant duration. Previous acoustic analyses and perceptual studies have also found that the spectral properties of the frication noise, the amplitude of the noise relative to the adjacent vowel(s), and the duration of the noise were distinctive attributes of fricatives and affricates. Behrens & Blumstein (1988) discuss the relevance of these cues to place of articulation. The duration cue is clearly associated with manner distinctions between stops, affricates, and

fricatives. Other parameters that have been shown to be important are the extent and duration of the formant transitions in adjacent vowels (Harris, 1958). These transitions were not measured with the implant speech processor, although they would have influenced the live-voice stimuli.

Finally, it should be noted that the high degree of redundancy within the two groups of parameters (rate, v/uv, F1, F2, A1) and (A2-A5) would probably be reduced for a wider range of phonemes and speakers. This may result in differences between the MSP1 and MSP2 processors for features other than place of articulation. In addition, background noise may affect some of the parameters and the redundant coding may help the user to perceive the true signal more easily under these conditions.

ACKNOWLEDGMENTS

Financial support was provided by the Australian Research Council and a Fogarty Senior International Fellowship # 1 FO6 TWO 1223-01. We would particularly like to thank the implant users who generously contributed their time and effort.

REFERENCES

Behrens, S. & Blumstein, S.E. (1988) *On the role of the amplitude of the fricative noise in the perception of place of articulation in voiceless fricative consonants,* J. Acoust. Soc. Am. 84, 861-867.

Blamey, P.J. Martin, L.F.A. & Clark, G.M. (1985) *A comparison of three speech coding strategies using an acoustic model of a cochlear implant,* J. Acoust. Soc. Am. 77, 209-217.

Blamey, P.J. Seligman, P.M. Dowell, R.C. & Clark, G.M. (1987a) *Acoustic parameters measured by a formant-based speech processor for a multiple-channel cochlear implant,* J. Acoust. Soc. Am. 82, 38-47.

Blamey, P.J. Dowell, R.C. Brown, A.M. Clark, G.M. & Seligman, P.M. (1987b) *Vowel and consonant recognition of cochlear implant patients using formant-estimating speech processors,* J. Acoust. Soc. Am. 82, 48-57.

Clark, G.M. Tong, Y.C. Patrick, J.F. Seligman, P.M. Crosby, P.A. Kuzma, J.A. & Money, D.K. (1984) *A multi-channel hearing prosthesis for profound-to-total hearing loss,* J. Med. Eng. Technol. 8, 3-8.

Dowell, R.C. Whitford, L.A. Seligman, P.M. Franz, B.K.-H. & Clark, G.M. (1990) *Preliminary results with a miniature speech processor for the 22 electrode Melbourne/Cochlear hearing prosthesis,* in Sacristan, T. Alvarez-Vicent, F. & Antoli-Candela, F. (Eds.) Proceedings of the XIV World Congress of Otorhinolaryngology, Head and Neck Surgery, (Kugler and Ghedini: Amsterdam).

Harris, K.S. (1958) *Cues for the discrimination of American English fricatives in spoken syllables,* Language & Speech 1, 1-7.

Schiffman, S.S. Reynolds, M.L. & Young, F.W. (1981) *Introduction to multidimensional scaling. Theory, methods, and applications,* (Academic: New York).

Seligman, P.M. (1987) *Speech processing strategies and their implementation,* Ann. Otol. Rhinol. Laryngol. Suppl. 128, 71-74.

Skinner, M.W. Holder, L.K. Holden, T.A. Dowell, R.C. Seligman, P.M. Brimacombe, J.A. & Beiter, A.L. (1991) *Performance for postlingually deaf adults with the wearable speech processor (WSPIII) and mini speech processor (MSP) of the Nucleus multi-electrode cochlear implant,* Ear Hear. 12, 3-22.

Strevens, P. (1960) *Spectra of fricative noise in human speech,* Language & Speech 3, 32-49.

Van Tasell, D.J. Soli, S.D. Kirby, V.M. & Widin, G.P. *Speech waveform envelope cues for consonant recognition,* J. Acoust. Soc. Am. 82, 1152-1161.